

# 基于机器学习和土壤关键要素的烤烟品质数字制图——以云南玉溪烟区为例<sup>①</sup>

陶 怡<sup>1,2</sup>, 王美艳<sup>1\*</sup>, 史学正<sup>1</sup>, 孙维侠<sup>1</sup>, 王世航<sup>2</sup>, 李湘伟<sup>3</sup>, 朱云聪<sup>3</sup>, 谢新乔<sup>3</sup>

(1 中国科学院南京土壤研究所, 南京 211135; 2 安徽理工大学空间信息与测绘工程学院, 安徽淮南 232001; 3 红塔烟草(集团)有限责任公司, 云南玉溪 653100)

**摘要:** 以中国典型烟区云南玉溪烟区为研究对象, 基于 156 个土壤-烤烟品质等级配对数据集, 通过主成分分析法筛选土壤关键要素, 应用反向神经网络(BPNN)、随机森林(RF)和支持向量机(SVM)3 种机器学习方法, 构建了基于土壤关键要素的烤烟品质等级预测模型, 实现了玉溪市烤烟品质等级空间预测与制图。研究结果表明: 基于 17 个土壤指标筛选出 11 个指标作为土壤关键要素, 其中土壤黏粒含量对烤烟品质等级的贡献率最大, 为 18.5%。独立验证结果显示, RF 模型的准确率和 Kappa 系数最高, 分别为 0.78 和 0.76, 预测效果最好, 其次是 SVM 模型, BPNN 模型最差。从召回率和精确率来看, RF 模型对烤烟品质等级正确分级效果的程度为五档>一档>二档。一档、五档烤烟主要集中分布在玉溪市东部, 最东部的华宁县是玉溪市最优质的烤烟作物产地。

**关键词:** 烤烟品质; 机器学习; 土壤关键要素; 空间分布; 制图

**中图分类号:** S151.9; S572; P283      **文献标志码:** A

## Digital Mapping of Flue-Cured Tobacco Quality Based on Machine Learning and Soil Key Elements—A Case Study of Yuxi Tobacco Area in Yunnan Province, China

TAO Yi<sup>1,2</sup>, WANG Meiyuan<sup>1\*</sup>, SHI Xuezhen<sup>1</sup>, SUN Weixia<sup>1</sup>, WANG Shihang<sup>2</sup>, LI Xiangwei<sup>3</sup>, ZHU Yuncong<sup>3</sup>, XIE Xinqiao<sup>3</sup>

(1 Institute of Soil Science, Chinese Academy of Sciences, Nanjing 211135, China; 2 School of Spatial Information and Surveying and Mapping Engineering, Anhui University of Science and Technology, Huainan, Anhui 232001, China; 3 Hongta Tobacco Group Company Limited, Yuxi, Yunnan 653100, China)

**Abstract:** In this study, Yuxi City of Yunnan Province, a typical tobacco-planting area in China, was selected as the study object, based on a dataset consisting of 156 pairs of soil-tobacco quality grades, soil key elements were identified through principal component analysis, and then three machine learning methods, namely the Back Propagation Neural Network (BPNN), Random Forest (RF) and Support Vector Machine (SVM) were employed to construct the prediction model of tobacco quality grade in order to achieve its spatial prediction and mapping. The results showed that based on 17 soil indicators, 11 specific indicators were identified as soil key elements, among these, clay content exhibited the highest contribution (accounting for 18.5%) to the variation in tobacco quality grades. The independent validation demonstrated that RF model achieved the highest accuracy (0.78) and Kappa coefficient (0.76) in the predictive performance, followed by SVM model, while BPNN model exhibited the least favorable outcomes. In terms of recall and precision, RF model demonstrated a descending level of accuracy in correctly categorizing tobacco quality grades, with the order of Level 5 > Level 1 > Level 2. Tobacco quality of Level 1 and 5 were predominantly distributed in the eastern part of Yuxi, with the easternmost Huanning County being the prime cultivation area for high-quality tobacco.

**Key words:** Tobacco quality; Machine learning; Soil key elements; Spatial distribution; Mapping

烤烟是一种重要的经济作物, 在世界各地广泛种植<sup>[1]</sup>。虽然烤烟品质受多种因素的影响, 但目前中国

典型优质烟区, 大多采用优良的烟草品种, 选择最适宜的气候区, 严格实施统一的施肥方案, 在这样的条

①基金项目: 红塔烟草(集团)有限责任公司科技项目(KY-Y60023015)资助。

\* 通讯作者(mywang@issas.ac.cn)

作者简介: 陶怡(1996—), 女, 安徽芜湖人, 硕士研究生, 研究方向为地理信息技术与应用。E-mail: 18855055211@163.com

件下, 土壤就成了目前影响烤烟品质最主要的因素。研究土壤关键要素如何影响烟叶品质, 对制定烤烟种植规划和提升烤烟品质均具有重要的指导意义。

土壤是烤烟养分的重要来源, 土壤要素丰缺直接影响烤烟生长发育的营养水平, 进而影响烟叶的品质<sup>[2]</sup>。胡玲等<sup>[3]</sup>通过估算土壤要素综合指标值对云南省植烟土壤要素状况进行综合性评价, 结果显示, 滇东和滇东南烟区的土壤速效磷含量有利于生产优质烤烟, 其余烟区速效磷含量偏高, 碱解氮、速效钾含量均值均达到丰富及以上水平; 张晓龙等<sup>[4]</sup>分析云南植烟土壤要素含量和烟叶品质之间的内在联系, 结果显示, 土壤要素含量对烟叶品质有直接效应, 烟叶中的钾主要来源于土壤钾素, 土壤中的氮素含量对烟叶总植物碱含量、总氮含量、浓度、刺激性等评价烟叶品质的指标有显著影响。目前有关土壤要素与烤烟品质的关联已进行了大量的研究工作<sup>[5-7]</sup>, 这些研究均表明烟叶品质与土壤要素有着密切的关系。

不同土壤要素会导致烟叶品质的不同, 同一烟草品种在不同土壤环境中的产质量也不相同<sup>[8]</sup>。目前, 烟草种植区划研究主要集中在基于生态条件的评估, 而基于烤烟品质的区划研究较少<sup>[9-10]</sup>。如, 李蒙等<sup>[9]</sup>以云南省普洱市为研究区, 基于普洱市多站点气象卫星资料, 运用隶属度函数模型, 与生态学相结合, 进行了烤烟栽培气候适宜性精细区划。这些研究由于缺少烤烟品质指标作为靶向数据, 往往受到主观经验的限制, 缺少一定的客观性。土壤对烤烟品质的影响是综合性的, 且这种影响并不是普通的线性关系, 很难用传统统计分析方法进行预测。机器学习能够处理复杂的非线性问题, 其通过分析处理多维数据, 去除冗余信息, 可进一步提取事物的本质规律及维度间的复杂联系<sup>[11]</sup>。反向神经网络(Back propagation neural network, BPNN)、随机森林(Random forest, RF)和支持向量机(Support vector machine, SVM)是机器学习中使用较多的3种模型, 在烤烟方面的应用主要有烤烟叶片叶绿素含量预测<sup>[12-13]</sup>、烤烟疾病检测<sup>[14-15]</sup>、烤烟烟叶分级<sup>[16-17]</sup>。如, 王韦燕等<sup>[12]</sup>将各特征波段下的光谱反射率作为模型的输入变量, 使用该3种机器学习模型对烤烟叶片叶绿素含量进行了预测。但基于土壤关键要素, 通过机器学习模型进行分析并以烤烟品质等级作为靶向预测的研究尚缺乏。为此, 本研究以玉溪烟区为研究对象, 评估了BPNN、RF、SVM 3种机器学习模型在烤烟品质等级预测中的精度, 并基于土壤关键要素和预测模型进行了烤烟品质等级的空间预测制图, 分析了烤烟等级的空间分异特征, 以

为优质烤烟收购、烟区优化布局及可持续发展提供的参考和支撑。

## 1 材料与方法

### 1.1 研究区概况

玉溪市( $23^{\circ}19' \text{N} \sim 24^{\circ}58' \text{N}$ ,  $101^{\circ}16' \text{E} \sim 103^{\circ}09' \text{E}$ )位于云南省中部(图 1), 北接昆明市, 西南连普洱市, 东南邻红河州, 西北靠楚雄州。年平均气温  $16.4 \sim 24.6^{\circ}\text{C}$ , 年日照时数  $2\,286.3 \sim 2\,085.3 \text{ h}$ , 年降水量  $674 \sim 1\,023 \text{ mm}$ 。其中大部分地区的海拔在  $1\,500 \sim 1\,800 \text{ m}$ , 地形以山地、高原、丘陵和岗地为主。玉溪烟区植烟面积达  $4 \times 10^4 \text{ hm}^2$ <sup>[18]</sup>, 主要植烟土壤类型为红壤、黄壤、紫色土和水稻土等<sup>[19-20]</sup>。

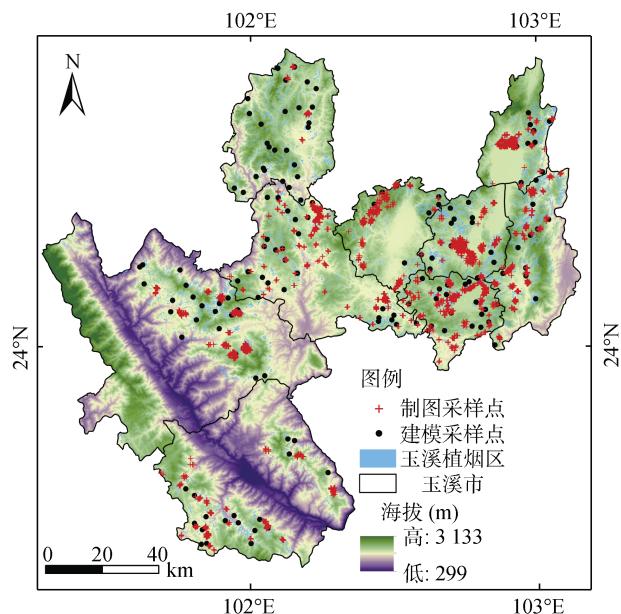


图 1 研究区域的位置和样本点的分布  
Fig. 1 Location of study area and sampling sites

### 1.2 样品采集与测定

2020年在玉溪植烟区布设典型烟田156个(图1), 在烤烟成熟采摘季, 进行土壤样品和烤烟样品的采集与制备。通过土壤样品分析测试与烤烟样品评吸, 获得土壤-烤烟品质等级的配对指标数据集, 用于模型构建。

土壤样品采集与分析: 在采样点田块, 通过5点梅花采样法采集耕层0~15 cm常规土壤混合分析样品1份, 采用100 cm<sup>3</sup>环刀采集原状样品3个。环刀样品用于土壤容重(BD)与土壤通气孔隙(PV)的测定, 混合样品用于常规指标测定, 包括土壤颗粒组成、pH、氯离子(Cl)、全氮(TN)、全钾(TK)、全磷(TP)、碱解氮(AN)、速效钾(AK)、有效磷(AP)、交换性钙

(Ca)、交换性镁(Mg)、有效硼(B)、有效锌(Zn)。其中, BD 采用环刀法(100 cm<sup>3</sup>) 测定; PV 采用砂箱法<sup>[21-22]</sup> 测定; 砂粒(Sand)、粉粒(Silt)和黏粒(Clay)含量采用吸管法(美国制)<sup>[23]</sup> 测定; pH、Cl、TN、TK、TP、AN、AK、AP、Ca、Mg、B、Zn 的具体测定方法参考鲁如坤<sup>[24]</sup>介绍的方法进行。

**烤烟样品采集与测定:** 采集标定样点田块烟叶, 由当地农户自行烘烤后, 采集 C3F(中橘三)等级的烟叶样品, 由红塔集团专业人员评吸。烟叶外观质量的指标包括颜色、成熟度、结构、身份、油分和色度; 感观质量的指标包括香气质、香气量、浓度、杂气、劲头、刺激性、余味、燃烧性和灰色。中国烟草总公司郑州烟草研究院(CNTC)的专家根据烟草行业感官评价标准(YC/T138—1998<sup>[25]</sup>)对烟叶进行外观和感官评吸, 并定性分为 5 个等级, 详细信息见表 1。

表 1 烤烟品质分级及数量统计  
Table 1 Tobacco quality grading and quantity statistics

烤烟品质等级	样本数量
一档(最好)	13
二档	16
三档	22
四档	18
五档(最差)	87

### 1.3 土壤关键要素遴选

通过主成分分析(Principal component analysis, PCA), 筛选影响烤烟品质等级的土壤关键要素, 构建最小指标数据集。主成分分析是一种应用广泛的特征降维方法, 其基本原理是将数据通过正交投影或者转换成低维子空间, 从而使数据投影方差最大化, 以少数不相关变量代替原始可能存在相关性的多维变量<sup>[26-27]</sup>。

### 1.4 数据处理与模型构建

**1.4.1 数据预处理** 为避免不同输入变量之间数量级别和量纲的影响, 降低异常数据对模型预测结果准确性的扰动, 将输入数据进行归一化处理, 计算公式为:

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

式中:  $x^*$  为归一化后的输入变量;  $x$  为原始输入变量;  $x_{\min}$  为原始输入变量最小值;  $x_{\max}$  为原始输入变量最大值。

**1.4.2 机器学习方法与模型构建** 基于 156 个土壤-烤烟品质等级配对样本数据集, 构建烤烟品质等级预测模型, 建模数据中随机选出 70% 作为训练集, 30% 作为测试集。选择 BPNN、RF 和 SVM 3 种机器学习方法, 以土壤关键要素作为预测因子, 以烤烟

品质等级作为分级结果, 通过参数调整, 构建分级预测模型。本研究用于机器学习模型的库来源于 Python 3.8.8 环境下调用的 Scikit-learn 包<sup>[28]</sup>。模型输入土壤指标 11 个, 预测烤烟品质等级 5 个。

1) BPNN。BPNN 是一种按照误差逆向传播算法训练的多层前馈神经网络, 其本身的学习规则是运用最速下降法不断调整网络的权值和阈值<sup>[29]</sup>。模型参数 input = 11, hidden = 8, output = 5, activation = ‘sigmoid’ 函数, learning\_rate = 0.01。

2) RF。RF 构造了多个决策树, 当需要对某个样本进行预测时, 统计森林中每棵树对该样本进行预测, 然后通过投票法从这些预测结果中选出最后的结果<sup>[30]</sup>。模型参数 n\_estimators = 50, max\_depth = 5, max\_features = 5, min\_samples\_split = 2。

另外, 该模型可以基于基尼指数量化评估输入变量中各指标的贡献程度, 增加 RF 模型的可解释性<sup>[31]</sup>, 计算公式为:

$$p_r = \frac{\sum_{a=1}^k \sum_{b=1}^t D_{G,rab}}{\sum_{r=1}^m \sum_{a=1}^k \sum_{b=1}^t D_{G,rab}} \quad (2)$$

式中:  $m$ 、 $k$ 、 $t$  分别为输入变量评价指标总数、分级树棵数和单棵树节点数;  $D_{G,rab}$  为第  $r$  个评价指标在第  $a$  棵树的第  $b$  个节点的基尼系数减少值;  $p_r$  为第  $r$  个评价指标在所有评价指标中的重要程度。

3) SVM。SVM 是一类按照监督学习方式对数据进行二元分类的广义线性分类器, 它可以通过核方法进行非线性分类<sup>[32]</sup>。模型参数 C=2, kernel = ‘linear’ 函数。

**1.4.3 模型预测性能评价** 准确率(Accuracy): 预测正确的结果占整个样本的百分比, 计算公式为:

$$\text{Accuracy} = \frac{tp + tn}{tp + tn + fp + fn} \quad (3)$$

精确率(Precision): 预测正确的正样本占所有预测为正样本的比例, 计算公式为:

$$\text{Precision} = \frac{tp}{tp + fp} \quad (4)$$

召回率(Recall): 样本中的正样本被预测正确的比例, 计算公式为:

$$\text{Recall} = \frac{tp}{tp + fn} \quad (5)$$

式中: tp 表示预测为正样本, 实际也是正样本; tn 表示预测为负样本, 实际也是负样本; fp 表示预测为正样本, 实际是负样本; fn 表示预测为负样本, 实际是正样本。

F1-分数(F1-score): 该评价标准结合了精确率与

召回率,能很好地反映模型的预测能力,计算公式为:

$$F1\text{-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$

期望一致率( $P_e$ ): 被正确预测的样本数占所有样本数( $n$ )的比例,计算公式为:

$$P_e = \frac{\sum_{i=1}^n a_i \times b_i}{n^2} \quad (7)$$

式中:  $i$  表示类别;  $a_i$  表示实际  $i$  类别的样本数;  $b_i$  表示预测的  $i$  类样本数。

Kappa 系数: 用于一致性检验的指标,也可以用于衡量分级的效果,取值为  $-1 \sim 1$ ,通常大于 0,计算公式为:

$$\text{Kappa} = \frac{\text{Accuracy} - P_e}{1 - P_e} \quad (8)$$

## 1.5 玉溪市烤烟品质等级空间预测与制图

进行烤烟品质等级的空间预测与制图,首先需要获取预测模型所需的空间数据,为此,收集整理玉溪植烟区近 3 年土壤监测数据,包括 2 401 个样点(图 1),使用 ArcGIS10.8 软件,通过普通克里格插值方法,得到玉溪市土壤关键要素的空间分布图,以此作为输入变量,放入调参好的 BPNN、RF、SVM 三个机器学习模型中,预测得到玉溪市的烤烟品质等级空间分布图。

数据统计和绘图采用 Excel 2016 和 Origin Pro 2021 软件, 主成分分析采用 SPSS 22.0 软件。

## 2 结果与分析

### 2.1 影响烤烟品质等级的土壤关键要素遴选

本研究采用主成分分析法从 17 个土壤物理和化

学指标中提取了 6 个主成分,第一、二、三、四、五、六主成分的贡献率分别为 22.86%、13.84%、13.17%、11.67%、8.45%、6.04%,累积贡献率达 76.03%(表 2)。从成分矩阵(表 3)中选取每个主成分中载荷值大于 50% 的指标,可以得到 PC1 主要由 Sand、Silt 和 Clay 决定,PC2 主要由 pH 和 Cl 决定,PC3 主要由 TK 和 AN 决定,PC4 主要由 AP 和 Ca 决定,PC5 主要由 Mg 决定,PC6 主要由 Zn 决定。因此,选取 Sand、Silt、Clay、pH、Cl、TK、AN、AP、Ca、Mg 和 Zn 共计 11 个土壤指标作为预测烤烟品质等级的土壤关键要素,即机器学习模型的输入变量。

表 2 总方差解释

Table 2 Explanation of total variance

主成分	初始特征值		
	总计	方差百分比(%)	累积贡献率(%)
1	4.12	22.86	22.86
2	2.49	13.84	36.70
3	2.37	13.17	49.87
4	2.10	11.67	61.54
5	1.52	8.45	69.99
6	1.09	6.04	76.03
7	0.73	4.06	80.09
8	0.60	3.80	83.89
9	0.52	3.26	87.15
10	0.40	2.82	89.97
11	0.39	2.57	92.54
12	0.31	1.92	94.46
13	0.30	1.88	96.34
14	0.21	1.65	97.99
15	0.14	1.02	99.01
16	0.06	0.98	99.99
17	0.00	0.01	100.00

表 3 主成分系数矩阵  
Table 3 Principal component coefficient matrix

指标	PC1	PC2	PC3	PC4	PC5	PC6
Sand	-0.51	0.50	0.04	-0.17	0.39	0.17
Silt	0.60	-0.43	0.09	0.50	-0.22	-0.01
Clay	0.56	-0.40	-0.16	-0.34	-0.36	-0.24
BD	0.03	-0.25	0.21	0.15	0.01	0.24
PV	0.41	-0.14	0.35	0.07	-0.48	0.16
pH	0.47	0.64	0.43	0.06	-0.03	0.02
Cl	0.46	0.62	0.03	0.08	-0.12	0.11
TN	0.39	-0.48	-0.15	0.14	0.47	0.19
TK	0.44	-0.15	0.70	-0.34	0.15	0.16
TP	0.42	-0.10	0.04	-0.37	0.09	0.17
AN	0.44	-0.03	-0.64	0.26	0.33	-0.10
AK	0.44	0.04	-0.38	-0.20	0.05	-0.15
AP	0.47	0.52	-0.36	-0.57	-0.20	0.04
Ca	0.33	0.04	0.49	0.52	0.20	-0.17
Mg	0.42	0.29	0.49	0.03	0.52	-0.18
B	0.24	0.48	-0.45	0.16	-0.30	0.25
Zn	0.48	0.47	-0.04	0.31	-0.03	-0.52

## 2.2 烤烟品质等级预测模型的精度评价

3种机器学习模型的精度评价结果见表4。从个体看，5个烤烟级别的评价指标数值之间存在差异，这表明3种模型不能完全正确地作用于同一年级，即同一档烤烟。从Recall和Precision看，RF模型能够正确分级一档、二档和五档烤烟，预测为一档、二档和五档的烤烟中分别有88%、71%和94%的烤烟实际上是一档、二档和五档，这说明RF模型对烤烟品质等级正确分级效果的程度为五档>一档>二档。从

F1-score看，RF模型对二档、四档、五档烤烟的预测得分范围为0.03~0.14，相距较近。BPNN模型和SVM模型对三档、四档烤烟的预测评价指标得分均为0，此外，BPNN模型对一档的也为0，这说明BPNN模型对一档、三档和四档烤烟的分级效果较差，SVM模型对三档和四档烤烟的分级效果较差。从整体上看，RF模型的Accuracy和Kappa系数最高，分别为0.78和0.76，RF模型的预测效果最好，其次是SVM模型，BPNN模型最差。

表4 三种预测模型的评价指标得分  
Table 4 Evaluation index scores of three prediction models

模型	烤烟品质等级	Precision	Recall	F1-score	Accuracy	Kappa
BPNN	一档	0	0	0	0.72	0.68
	二档	0.29	1.00	0.44		
	三档	0	0	0		
	四档	0	0	0		
	五档	0.80	1.00	0.89		
RF	一档	0.88	1.00	0.67	0.78	0.76
	二档	0.71	1.00	0.83		
	三档	0.67	0.50	0.57		
	四档	1.00	0.75	0.86		
	五档	0.94	1.00	0.97		
SVM	一档	0.67	0.67	0.67	0.76	0.74
	二档	0.71	0.83	0.77		
	三档	0	0	0		
	四档	0	0	0		
	五档	0.83	1.00	0.91		

## 2.3 土壤关键要素对烤烟品质等级的贡献率

RF模型在本研究中的制图精度最高，RF模型最大的优点是在模型训练的同时可以评估输入变量的贡献率，如图2所示。土壤颗粒组成的累积贡献率达44.2%，说明土壤质地对烤烟品质起决定性的作用。土壤质地是决定土壤蓄水、保温和耕性等的重要因素，因此，不同质地土壤在农业生产性状上表现有很大差异。其中，土壤黏粒的贡献率最大，为18.5%；砂粒的贡献率次之，为16%；粉粒的贡献率最小，为9.7%。土壤关键要素对玉溪烟区烤烟品质等级的重要性排序：Clay>Sand>Silt>Mg=Ca>AN>pH=Cl>TK>AP>Zn。

## 2.4 基于三种机器学习模型预测的玉溪烟区烤烟品质空间分布

应用3种机器学习方法，基于土壤关键要素的烤烟品质等级预测与制图，如图3所示。整体来看，基于3种机器学习模型的玉溪烟区烤烟品质等级预测

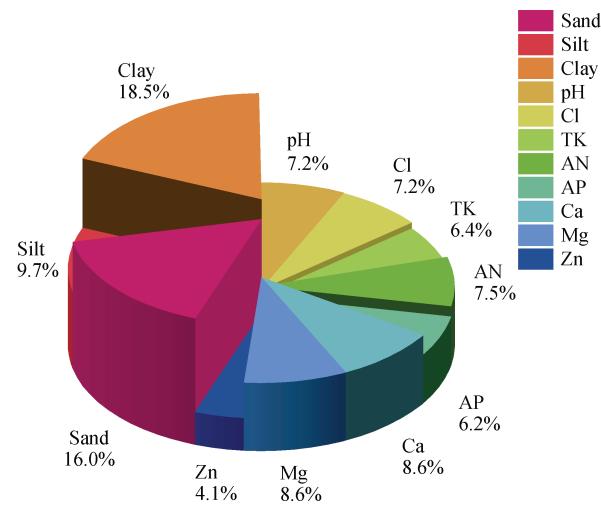


图2 土壤关键要素对烤烟品质等级的贡献  
Fig. 2 Contributions of key soil elements to tobacco quality grade

结果都展现出了相似的空间分布格局，特别是基于RF模型生成的分布图与SVM模型比较相似。一档、五档烤烟主要集中分布在玉溪市东部，二档主要分布

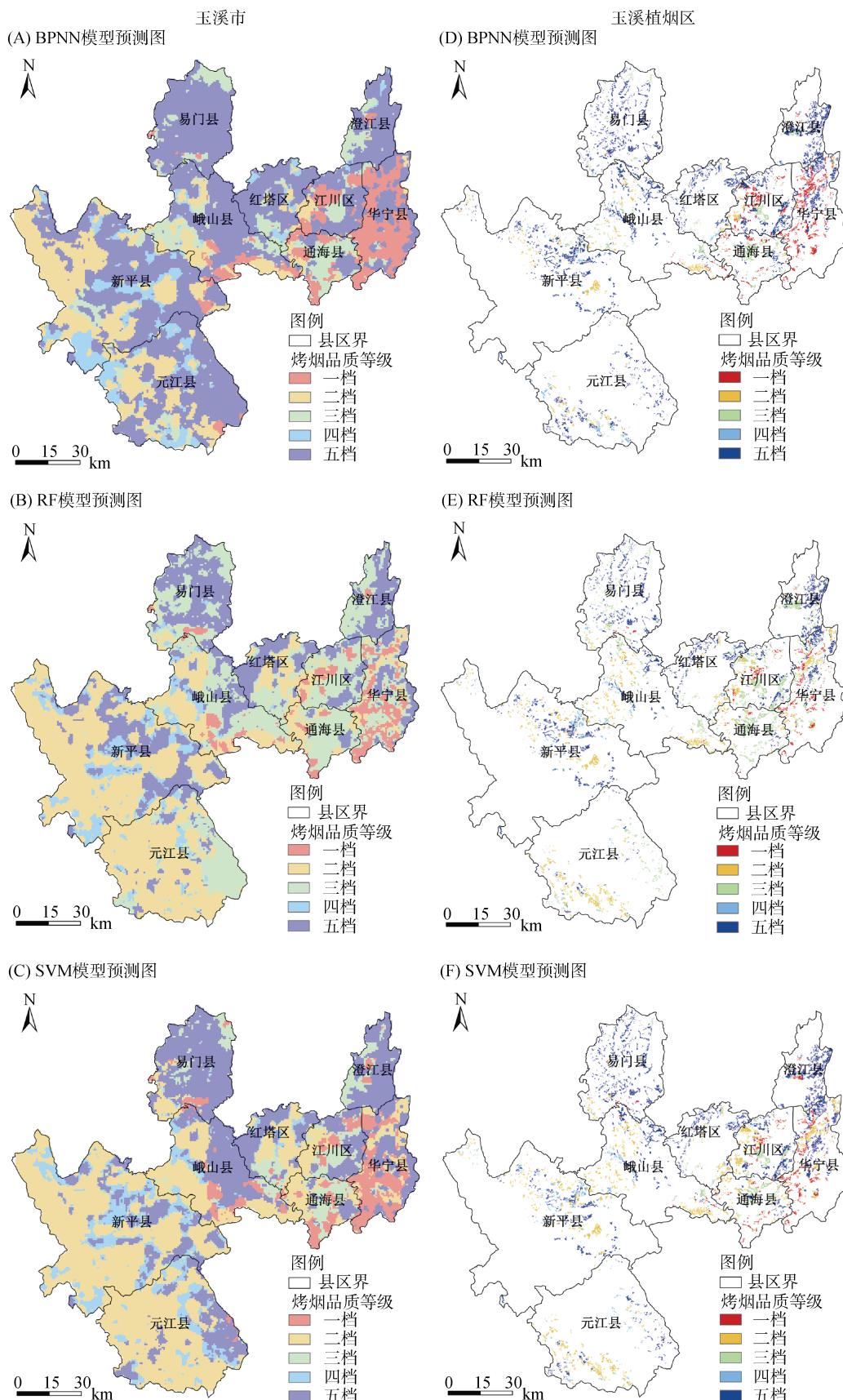


图 3 基于机器学习模型预测的玉溪市(A, B, C)和玉溪植烟区(D, E, F)烤烟品质分布

Fig. 3 Distribution of tobacco quality grade based on machine learning model prediction for whole Yuxi City (A, B, C) and tobacco-planting area (D, E, F)

在玉溪市西南部，三档主要分布在玉溪市中部。具体来看，基于 RF 模型和 SVM 模型的一档和二档烤烟面积均大于 BPNN 模型，且五档烤烟面积均小于 BPNN 模型。大部分一档烤烟集中分布在玉溪市东部，即峨山县东南部、通海县南部等地区；五档烤烟也集中分布在玉溪市东部，即华宁县北部、澄江县东南部等地区。从 RF 模型具体结果来看，二档烤烟在玉溪市元江县、新平县及峨山县大面积覆盖，然而在玉溪实际植烟区这 3 个县的覆盖较少，特别是元江县和新平县除东部以外的地区，因此建议可以考虑适当扩大植烟区范围，大面积种植二档烤烟。

从烤烟品质分级的结果看，基于 RF 模型预测的

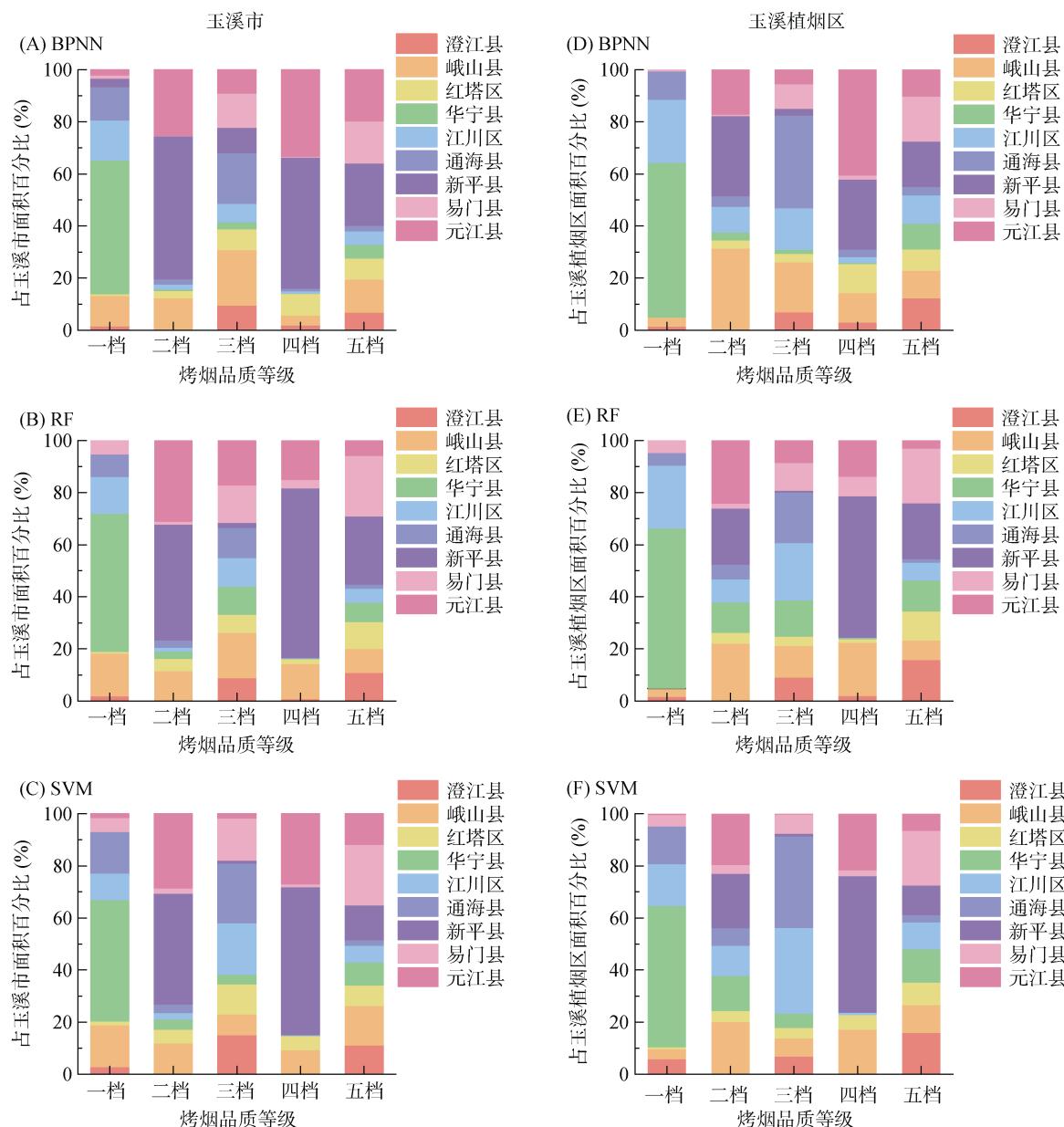


图 4 预测的烤烟品质各等级中玉溪市(A, B, C)和玉溪植烟区(D, E, F)各县区所占面积的百分比  
Fig. 4 Predicted area percentages of tobacco quality levels in whole Yuxi City (A, B, C) and tobacco-planting regions (D, E, F)

玉溪市和玉溪植烟区的一档烤烟中华宁县所占面积的百分比最高，如图 4 所示，分别为 53.08% 和 61.46%，表明基于 RF 模型预测的玉溪植烟区的一档烤烟中华宁县所占面积比玉溪市高。华宁县植烟区面积只占整个华宁县的 21.71%，没有达到 50%。华宁县是玉溪市非常适合优质烤烟种植和生产的县区，其土壤性状优越，但植烟面积却不大，除现有的植烟区以外还有其他适合种植优质烤烟的植烟区，可扩大植烟区面积以满足更多优质烤烟的生长需求。从其他级别看，二档和四档烤烟中新平县所占面积的百分比最大，三档和五档烤烟中各县区占比较平均，没有占比很明显的县区。

从玉溪市各县区看, 基于 RF 模型预测的玉溪市和玉溪植烟区的华宁县中一档烤烟所占面积的百分比最高, 澄江县、红塔区、易门县中五档烤烟所占面积的百分比最高, 如图 5 所示。此外, 易门县中五档烤烟所占面积的百分比又是其他各县区中各档次烤

烟占比最多的, 玉溪市为 54.89%, 玉溪植烟区为 60.65%。由此说明, 易门县的烤烟品质较差, 由于土壤性状和烤烟品质关系密切, 其土壤性状可能也较差, 可以考虑对易门县的土壤进行科学施肥, 改良土壤性状。

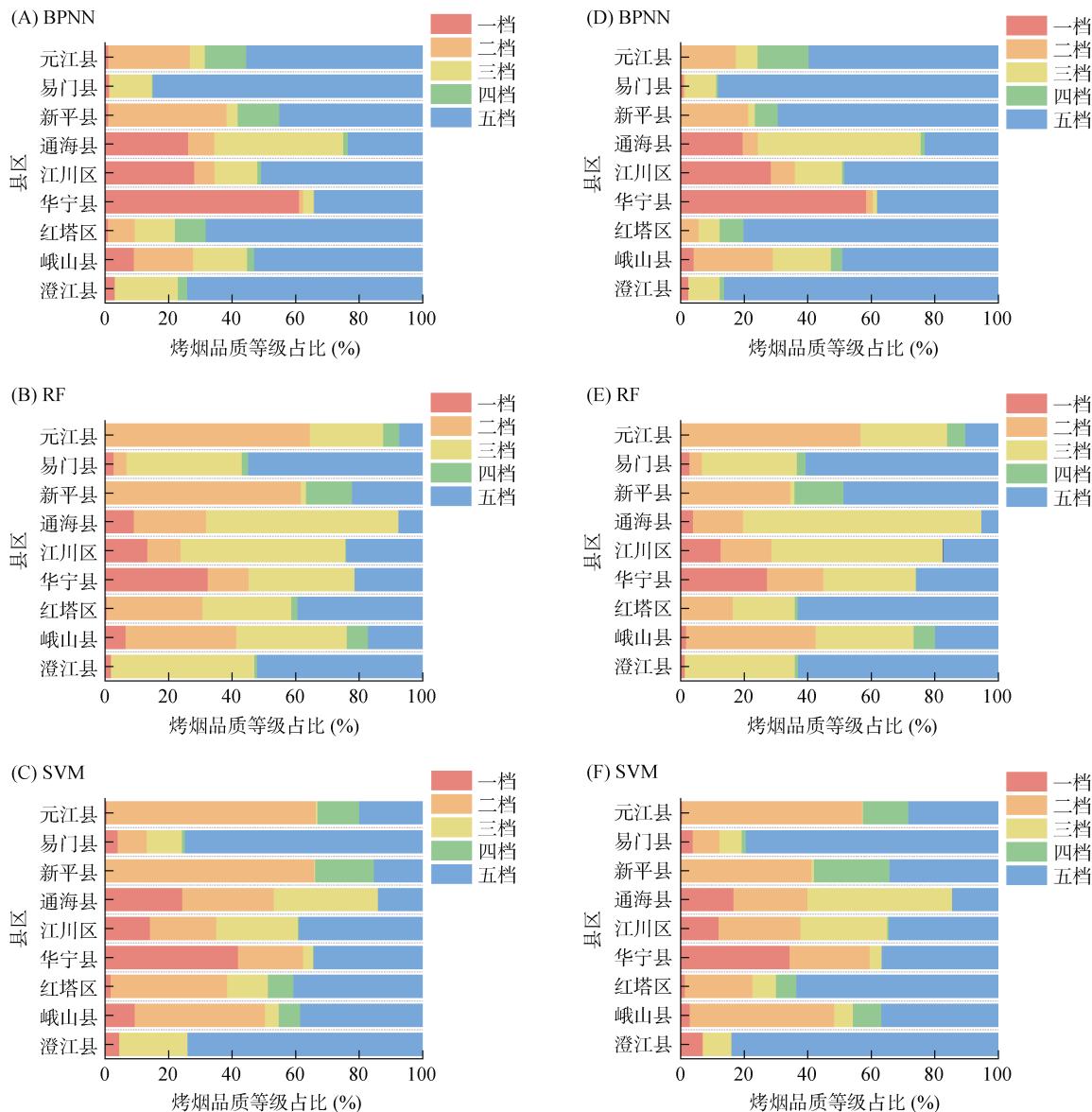


图 5 预测的玉溪市(A, B, C)和玉溪植烟区(D, E, F)各县区中各烤烟品质等级所占面积的百分比  
Fig. 5 Predicted area percentages of tobacco quality levels in whole Yuxi City (A, B, C) and tobacco-planting regions (D, E, F)

### 3 讨论

本研究结果表明, RF 模型预测效果表现最佳, 最适用于玉溪烟区烤烟品质预测与制图, 与徐佳等<sup>[33]</sup>的研究结果较为一致, 其以我国东部 4 省 2 市(北京、天津、河北、山东、安徽和江苏)为研究区, 基于 746 个土壤表层样本的理化性质和光谱数据, 使用 BPNN 和 RF 两种机器学习模型对海拔高度、年均温、年均

降雨量和地表温度 4 个关键环境要素进行预测, 研究结果表明, 相较于 BPNN 模型, RF 模型的预测效果更好。RF 模型能更好地刻画土壤与烤烟品质之间复杂的关系, 其原因可能是由于随机抽取的训练样本难以泛化到整个研究区的真实环境数据集上, 而 BPNN 模型是网络状模型, 模拟大脑皮层神经元结构, 网络拟合准确度高, 学习过多的特殊样本, 容易使模型对其他样本的反应有偏差, 造成过拟合<sup>[34]</sup>; RF 模型是

集合式树状模型,融合了大量决策树进行训练,可以减轻过拟合问题<sup>[35]</sup>。RF模型的预测是每个单独树预测的集合,能够充分表达土壤和烤烟品质等级的关系。RF模型利用了这种优势,在更新烤烟品质等级图时做出了更精确的预测。此外,在五元分级的情况下,SVM模型使用超平面将五级分开,如果两个烤烟品质等级的空间分布较为分散,分类超平面将难以区分两个烤烟品质等级,在推理过程中容易出现烤烟品质等级的位移和破碎化<sup>[36]</sup>。因此,SVM模型和BPNN模型无法达到与RF模型相同的预测烤烟品质等级分布精度的水平。

本研究中,土壤关键要素对玉溪烟区烤烟品质等级的重要性排序:黏粒>砂粒>粉粒>交换性镁=交换性钙>碱解氮>土壤pH=氯离子>全钾>全磷>有效锌。贺丹锋等<sup>[37]</sup>以云南罗平烟区10个乡镇中部烟叶和对应土样样品为研究对象,运用主成分分析法计算得到各参评指标的权重值:总氮>钾氯比>还原糖>总糖>糖碱比>烟碱>氯>钾>氮碱比;龚玖零<sup>[10]</sup>以云南保山市62个植烟乡镇的土壤、烟叶为研究对象,运用主成分分析法确定了各土壤养分指标的权重:水溶性氯离子>碱解氮>速效钾>有机质>有效硼>有效锌>速效磷>土壤pH。由此可以看出,土壤关键要素的重要性没有很清晰的排序,针对不同的研究区,土壤属性的重要性不同。土壤质地是影响烤烟矿质营养吸收与累积的重要环境因素<sup>[38]</sup>。窦逢科等<sup>[39]</sup>研究表明,土壤表层质地为砂壤土,亚表层质地为较黏且排水性良好的土层更易产出高质量烟叶。亢贵霞等<sup>[40]</sup>认为砂土抗旱能力弱,保肥性能差;黏土通气性差,有机质含量丰富;壤土兼有两者的优点,是理想的植烟土壤。这些研究都论证了土壤质地在很大程度上影响烟叶品质。本研究也表明,土壤质地对烤烟品质累积贡献率达44.2%,土壤质地对烤烟品质起决定性的作用。土壤质地中土壤黏粒的贡献率最大,达到了18.5%,证明玉溪烟区土壤质地较黏,因此,建议可以考虑增施有机肥。高黏粒含量土壤增加有机肥能促进养分的吸收和固定,有利于颗粒间形成大团聚体,土壤碳汇功能、水热性能改善,从而改善烟叶化学成分协调性<sup>[41]</sup>。

玉溪市的成土环境差异与人类活动的剧烈影响造成该区域土壤属性的独特空间变异性,导致土壤中生长的烤烟品质差异明显。如东部地区具有良好的农用地资源,东部典型植烟区华宁县青龙镇、宁州镇等土壤质地均以黏壤土为主,壤土次之,适宜烟叶生长<sup>[19]</sup>。优质植烟土壤的全氮、全磷、全钾和

速效氮皆处于丰富水平,速效磷和速效钾处于适中水平,土壤有机质含量丰富,pH均值为6.1,处于5.5~6.5的适宜范围内,供氮能力强,可满足烤烟生长养分需求<sup>[42]</sup>。玉溪烟区植烟土壤面积较大,且区域水系密集,有抚仙湖、星云湖、杞麓湖等湖泊,灌溉条件相对优越,长期连作以及农田肥料投入量较高,耕作历史悠久,造成玉溪市烤烟品质等级分异明显<sup>[43]</sup>。易门县高程标准差偏大,地形波动较大,易受低温冷害、干旱等灾害的影响,烤烟品质和产量呈现下降趋势<sup>[44]</sup>。

由于所收集资料和数据的限制,本研究仅考虑了土壤对玉溪烟区烤烟品质的影响,温度、降水、光照、地形等自然条件也是影响烤烟品质的重要因素,今后的研究可综合考虑这些条件并纳入评价因子,提高预测模型的精度,为制定烤烟种植规划和提升烤烟品质提供更准确的技术依据和参考。

## 4 结论

1) 土壤颗粒组成和养分含量等11个指标可以作为预测玉溪烟区烤烟品质的土壤关键要素,其中黏粒含量对烤烟品质等级的贡献率最大。

2) 独立验证结果显示,RF模型预测效果最好,其次是SVM模型,BPNN模型最差。RF模型适用于玉溪烟区的烤烟品质等级空间分布预测。

3) 基于3种机器学习模型的玉溪烟区烤烟品质等级预测结果均展现了相似的空间分布格局,一档、五档烤烟主要集中分布在东部,二档主要分布在西南部,三档主要分布在中部。最东部的华宁县是玉溪市最优质的烤烟产地,其一档烤烟面积占比在所有县区中最大,可适当扩大该县的植烟面积。

**致谢:**感谢玉溪烟区相关人员在调查采样过程中给予的支持和帮助。

## 参考文献:

- [1] Yang L Y, Wang L T, Ma J H, et al. Effects of light quality on growth and development, photosynthetic characteristics and content of carbohydrates in tobacco (*Nicotiana tabacum* L.) plants[J]. *Photosynthetica*, 2017, 55(3): 467–477.
- [2] 郑立臣, 宇万太, 马强, 等. 农田土壤肥力综合评价研究进展[J]. 生态学杂志, 2004, 23(5): 156–161.
- [3] 胡玲, 周丽娟, 王娟, 等. 云南烟区植烟土壤养分状况综合评价[J]. 河南农业科学, 2014, 43(7): 52–59.
- [4] 张晓龙, 薛红芬, 罗华元, 等. 云南植烟土壤养分含量、微生物数量、烟叶品质的相互关系[J]. 河南农业科学, 2015, 44(6): 68–71.

- [5] 方秀, 王林, 卢秀萍, 等. 曲靖烟区土壤磷锌互作对烟叶磷锌含量和品质的影响[J]. 土壤, 2018, 50(5): 894–901.
- [6] 潘金华, 王美艳, 史学正, 等. 玉溪烟区土壤钾镁交互作用对烤烟化学及感官品质的影响[J]. 土壤, 2022, 54(3): 490–497.
- [7] 蔡寒玉, 廖文程, 李兰周, 等. 云南丽江植烟土壤养分状况综合评价[J]. 云南农业大学学报(自然科学), 2016, 31(2): 341–347.
- [8] 陈江华, 刘建利, 李志宏, 等. 中国植烟土壤及烟草养分综合管理[M]. 北京: 科学出版社, 2008.
- [9] 李蒙, 杨明, 王伟, 等. 云南普洱市烤烟种植气候适宜性精细化区划[J]. 作物杂志, 2010(6): 75–79.
- [10] 龚玖零. 基于植烟土壤养分及烟叶质量的云南保山市植烟分区[D]. 昆明: 云南农业大学, 2015.
- [11] Wan A P, Jie Y, Ting C, et al. Dynamic pollution emission prediction method of a combined heat and power system based on the hybrid CNN-LSTM model and attention mechanism[J]. Environmental Science and Pollution Research International, 2022, 29(46): 69918–69931.
- [12] 王韦燕, 冯文强, 常乃杰, 等. 基于光谱预处理和机器学习算法的烤烟叶绿素含量预测[J]. 中国土壤与肥料, 2023(3): 194–201.
- [13] 赖佳政, 李贝贝, 程翔, 等. 基于无人机高光谱遥感的烤烟叶片叶绿素含量估测[J]. 智慧农业(中英文), 2023, 5(2): 68–81.
- [14] Zhu H Y, Chu B Q, Zhang C, et al. Hyperspectral imaging for presymptomatic detection of tobacco disease with successive projections algorithm and machine-learning classifiers[J]. Scientific Reports, 2017, 7(1): 4125.
- [15] Gu Q, Sheng L, Zhang T H, et al. Early detection of tomato spotted wilt virus infection in tobacco using the hyperspectral imaging technique and machine learning algorithms[J]. Computers and Electronics in Agriculture, 2019, 167: 105066.
- [16] 雒慧心. 基于深度学习的烤烟分选算法研究[D]. 北京: 北京交通大学, 2019.
- [17] 鲁梦瑶, 陈栋, 周强, 等. 基于深度学习的烟叶等级分类及特征可视化[J]. 烟草科技, 2023, 56(6): 92–100.
- [18] 王秀珍, 景元书, 谢新乔, 等. 玉溪烤烟低温冷害风险评价与区划[J]. 河南农业大学学报, 2020, 54(4): 704–710.
- [19] 田育天, 李湘伟, 谢新乔, 等. 云南典型植烟土壤通气孔隙及其主控因素研究[J]. 土壤学报, 2020, 57(6): 1430–1438.
- [20] 田育天, 李湘伟, 谢新乔, 等. 粿秆还田对云南典型烟区土壤物理性状的影响[J]. 土壤, 2019, 51(5): 964–969.
- [21] Moret D, Arrué J L. Dynamics of soil hydraulic properties during fallow as affected by tillage[J]. Soil and Tillage Research, 2007, 96(1/2): 103–113.
- [22] 王红兰, 唐翔宇, 鲜青松, 等. 紫色土水分特征曲线室内测定方法的对比[J]. 水科学进展, 2016, 27(2): 240–248.
- [23] 林业部. 森林土壤颗粒组成(机械组成)的测定: GB 7845—1987[S]. 北京: 中国标准出版社, 1987.
- [24] 鲁如坤. 土壤农业化学分析方法[M]. 北京: 中国农业科技出版社, 2000.
- [25] 国家烟草专卖局. 烟草及烟草制品 感官评价方法: YC/T 138—1998[S]. 北京: 中国标准出版社, 1998.
- [26] Hargrove L J, Li G L, Englehart K B, et al. Principal components analysis preprocessing for improved classification accuracies in pattern-recognition-based myoelectric control[J]. IEEE Transactions on Bio-Medical Engineering, 2009, 56(5): 1407–1414.
- [27] 杨博雄, 杨雨绮. 利用 PCA 进行深度学习图像特征提取后的降维研究[J]. 计算机系统应用, 2019, 28(1): 279–283.
- [28] Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in python[J]. The Journal of Machine Learning Research, 2011, 12: 2825–2830.
- [29] Liu Y T. Incomplete big data imputation mining algorithm based on BP neural network[J]. Journal of Intelligent & Fuzzy Systems, 2019, 37(4): 4457–4466.
- [30] Liu F, Wu H Y, Zhao Y G, et al. Mapping high resolution National Soil Information Grids of China[J]. Science Bulletin, 2022, 67(3): 328–340.
- [31] 吴孝情, 赖成光, 陈晓宏, 等. 基于随机森林权重的滑坡危险性评价: 以东江流域为例[J]. 自然灾害学报, 2017, 26(5): 119–129.
- [32] Cortes C, Vapnik V. Support-vector networks[J]. Machine Learning, 1995, 20(3): 273–297.
- [33] 徐佳, 刘峰, 吴华勇, 等. 基于人工神经网络和随机森林学习模型从土壤属性推测关键成土环境要素的研究[J]. 土壤通报, 2021, 52(2): 269–278.
- [34] Zhao H, Wang Y, Song J H, et al. The pollutant concentration prediction model of NNP-BPNN based on the INI algorithm, AW method and neighbor-PCA[J]. Journal of Ambient Intelligence and Humanized Computing, 2019, 10(8): 3059–3065.
- [35] Lindner C, Bromiley P A, Ionita M C, et al. Robust and accurate shape model matching using random forest regression-voting[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1862–1874.
- [36] Liu X Q, Zhu A X, Yang L, et al. Influence of legacy soil map accuracy on soil map updating with data mining methods[J]. Geoderma, 2022, 416: 115802.
- [37] 贺丹锋, 周冀衡, 孙书斌, 等. 罗平烟区不同植烟土壤养分状况对中部烟叶化学成分可用性的影响[J]. 天津农业科学, 2016, 22(5): 27–33.
- [38] 王雪婧. 基于 GIS 郏县烟草生态区划研究[D]. 郑州: 河南农业大学, 2013.
- [39] 窦逢科, 张景略. 烟草品质与土壤肥料[M]. 郑州: 河南科学技术出版社, 1992.
- [40] 亢贵霞, 徐宇航, 王玉川, 等. 毕节市植烟土壤分析与区划研究[J]. 安徽农业科学, 2013, 41(20): 8496–8500.
- [41] 葛楠楠, 石芸, 杨宪龙, 等. 黄土高原不同土壤质地农田土壤碳、氮、磷及团聚体分布特征[J]. 应用生态学报, 2017, 28(5): 1626–1632.
- [42] 陆俊平, 谢新乔, 李湘伟, 等. 玉溪烟区土壤主要理化性状与烟叶品质的相关性分析[J]. 黑龙江农业科学, 2022(10): 38–44, 49.
- [43] 谢新乔, 陆俊平, 田育天, 等. 玉溪市 100 m 级植烟土壤质地品质的区划研究[J]. 土壤学报, 2023, 60(4): 1113–1122.
- [44] 王秀珍. 玉溪烤烟气象灾害风险及作物模型适用性研究[D]. 南京: 南京信息工程大学, 2021.